

## PDF hosted at the Radboud Repository of the Radboud University Nijmegen

The following full text is a publisher's version.

For additional information about this publication click this link.

<http://hdl.handle.net/2066/46957>

Please be advised that this information was generated on 2017-12-06 and may be subject to change.

**NiCE Working Paper 07-107**

**November 2007**

# **The Empirical Dimensions of the Scale Effects Problem**

**André van Hoorn**

Nijmegen Center for Economics (NiCE)  
Institute for Management Research  
Radboud University Nijmegen

P.O. Box 9108, 6500 HK Nijmegen, The Netherlands

<http://www.ru.nl/nice/workingpapers>

## **Abstract**

Empirical work has refuted the prediction of new or endogenous growth theory that growth exhibits a scale effect: the rate of growth has not accelerated and larger countries do not grow faster than smaller countries do. The theoretical dimensions of this ‘scale effects problem’ have been explored at length, but this has not yet led to a satisfactory solution. Concerning its empirical dimensions, it has been suggested that problems of measurement, in particular the ‘quality-improvement problem’, causing price increases to be overstated and real growth to be understated can account for the absence of the predicted growth scale effects. We offer the first test of this possible explanation and find that these biases do not offer a solution to the scale effects problem. The upward bias in measured prices has not worsened and our empirical analysis dismisses the increasing importance of ‘unmeasurable’ sectors as a possible explanation. It shows that the bias due to problems of measurement needs to be implausibly high for growth to exhibit a scale effect. Since neither theoretical nor empirical work seems able to come up with a satisfactory solution to the problem of scale effects, it might be much more fundamental than often realised.

Correspondence: André van Hoorn, Radboud University Nijmegen, Department of Economics, PO Box 9108, 6500 HK, Nijmegen, the Netherlands, Tel: +31 243 612 344, Fax: +31 243 612 379, E-mail: [A.vanHoorn@fm.ru.nl](mailto:A.vanHoorn@fm.ru.nl).

Helpful comments from seminar participants of the International Network for Economic Research (INFER) Annual Conference and at Radboud University Nijmegen are gratefully acknowledged. Sjoerd Beugelsdijk provided especially useful suggestions.

## 1. Introduction

New or endogenous growth theory has greatly advanced our understanding of growth and development. Notably, it extends old, neoclassical growth theory, in which technological change is exogenous, by incorporating the notion that firms invest in the creation of new knowledge and that there are incentives for them to do so, notably monopoly profits (see Romer, 1994). A common feature of the early endogenous growth models (e.g. Romer, 1990; Aghion and Howitt, 1992) is that growth rates rise with scale and that growth accelerates. This theoretical prediction, however, is inconsistent with observed patterns of per-capita income, and growth scholars wrestle with this ‘problem of scale effects’ (Jones, 1995b: 761).

Several theoretical solutions to this scale effects problem have been proposed (amongst others, Smulders and van de Klundert, 1995; Segerstrom, 1998; Young, 1998) but these have only exacerbated the problem. As, for example, Li (2000, 2002) shows, only under very strict (knife-edge) assumptions is growth endogenous in these models; in fact, modelling (R&D-based) growth without scale effects generally implies a return to the neoclassical model in which exogenous factors determine long-run growth.<sup>1</sup>

The empirical work that has refuted the predicted growth scale effects is however not without frailties. Specifically, it has neglected important measurement problems that plague data on economic growth. In this paper, we aim to contribute to this gap and make some additional steps in the analysis of the measurement dimension of the scale effects problem. This is important because if measurement problems are the cause of our failure to establish accelerating growth, the scale effects prediction of endogenous growth models is not as problematic as sometimes suggested. If, on the other hand, accelerating

---

<sup>1</sup> The focus of this note is on R&D-based models of endogenous growth such as Romer (1990) or Aghion and Howitt (1992). Though different varieties of models in which growth is endogenously determined as the result of investments in knowledge or R&D exist, the prediction of growth scale effects is general; it is a feature also of, for example, AK models of growth such as Romer (1986) and Lucas (1988) (see Aghion and Howitt, 1998).

growth cannot be established on empirical grounds, a tremendous challenge arises to align endogenous growth theory with this empirical reality. Aghion and Howitt (1998: 438) identify the so-called ‘quality-improvement problem’ as the most serious measurement problem in this context. Following their lead, we focus on the role of prices, in particular the consumer price index (CPI), the upward bias therein, and the associated undermeasurement of real GDP growth to shed more light on this important issue.

Using data on real gross domestic product (GDP), number of hours worked and the number of employees, we test under what conditions accelerated growth occurs. Our analysis of a sample of nine countries and ten sectors suggests that measurement problems cannot account for the absence of the predicted growth scale effects. Firstly, we dismiss the possibility that the bias in calculating the CPI has worsened. If this were the case, this would account for non-accelerating growth. Tracing the development of the CPI bias over time shows that the upward bias has not increased. Secondly, using revised CPI figures, we subsequently correct for undermeasurement in our data on GDP per hour worked. We find no sign of an upward trend in productivity growth. Growth only exhibits a scale effect when we apply very large corrections, which are not in proportion to even the highest estimates for the actual bias in measured prices. Applying a range of robustness tests, we still find no evidence of accelerating growth in the productivity statistics.

Our analysis complements existing literature in two ways. First, our results suggest that accelerating growth cannot be established on empirical grounds. Second, we explore the implications of this conclusion for growth theory and the scale effects problem in specific. The remainder of this paper is organised as follows. We briefly discuss the theoretical origins and main empirical evidence on growth scale effects in Section 2. Section 3 discusses measurement errors in prices (i.e. in the CPI) and their relation to the scale effects problem. We test the prediction of scale effects / accelerating growth using data corrected for mismeasurement in Section 4. Finally, we discuss the broader implications of our findings in Section 5.

## 2. Scale Effects and Accelerating Growth

The possible effects of economy size or scale have long been an issue in economics. Petty (1682), for instance, already discusses the impact of population size on ‘convenient, commodious, and comfortable livings’. The more recent empirical work investigating the importance of economy size or scale on its rate of expansion follows the development of endogenous growth theory. Early R&D-based models such as the ones developed by Romer (1990) and Aghion and Howitt (1992), all predict a rise in growth rates as scale increases.

Before turning to the empirical aspects of the growth-scale problem, we provide a succinct discussion of its theoretical core. Refraining from a full formal discussion of the microfoundations of endogenous growth models, three simple equations capture the essence of these models (see, for example, Jones, 2003). Output ( $Y$ ) is produced using a fraction of the total labour force ( $L_Y$ ) and the stock of knowledge ( $A$ ):

$$Y = A^\sigma L_Y . \quad (1)$$

This production function exhibits constant returns to scale in the rivalrous input, labour, and increasing returns to scale in labour and the nonrivalrous input, knowledge, combined. New knowledge is created using a fraction of the total labour force ( $L_A$ ) and the existing stock of knowledge:

$$\dot{A} = \delta L_A A , \quad (2)$$

where  $\delta$  is a parameter determining the arrival rate of new knowledge. An important aspect of Equation 2 is that past R&D efforts contribute to the productivity of current R&D: there are intertemporal knowledge spillovers in the production of new knowledge. From combining Equation 1 and 2, it is easy to obtain an expression for the growth rate of output per worker ( $g_Y$ ):

$$g_y \equiv \frac{\dot{Y}}{Y} - \frac{\dot{L}}{L} = \sigma \frac{\dot{A}}{A} = \sigma g_A = \sigma \delta s L, \quad (3)$$

where  $0 < s < 1$  is the fraction of the total labour force active in the creation of new knowledge ( $L_A = sL$ ;  $L_Y = (1-s)L$ ). This last equation presents a suitable measure of scale (cf. Backus et al. 1992: 378): an economy is larger than another economy if it devotes a larger absolute amount of resources, e.g. labour, to the production of knowledge (R&D) than the other economy, *ceteris paribus*. Based on the above theoretical derivation, scholars have tried to empirically test endogenous growth theory's predictions, more specifically the growth scale effect.

The most influential study on growth scale effects is by Jones (1995a). Examining data for the United States and fourteen additional advanced OECD countries, he finds that growth in these countries has been remarkably stable, especially since the 1950s. Moreover, acknowledging that the scale of these economies has risen over time, the rates at which output grows do not show a persistent rise.

[Insert Figure 1 about here]

Figure 1 summarises some of Jones' evidence. It shows that, except for breaks around the Great Crash and World War II, growth of per-capita GDP for the U.S. and the G4 (the U.S. and France, Germany and Japan) has been steady since before 1890 and shows no sign of acceleration. TFP (total factor productivity) growth appears trendless as well, despite sustained augmentation of the number of scientists and engineers engaged in R&D. Other studies confirm that trends in output do not accelerate and that larger economies do not grow faster than smaller economies do (see, e.g., the review by Dinopoulos and Thompson, 1999). This leads Jones to conclude that the empirical evidence refutes the prediction of growth scale effects of endogenous growth models.

Based on these empirical studies, growth scholars have turned to the formal R&D-based growth models and tried to adapt these accordingly. As already mentioned, the proposed theoretical solutions to the scale effects problem are not without limitations, however. Growth models without scale effects resemble old, neoclassical models by

making long run growth exogenous again (Li, 2000, 2002). Meanwhile, the possibility that the measurement difficulties associated with the empirical work could somehow account for the absence of growth scale effects has received limited attention. In their textbook on endogenous growth, Aghion and Howitt (1998) identify several measurement problems that matter for endogenous growth theory. The most serious of these—that is, the problem with the biggest impact on measured growth—is the so-called ‘quality-improvement problem’ (438): ‘... to the extent that knowledge creation within business firms results in improved goods and services, the practical difficulties of dealing with new goods and quality improvements in constructing price indices imply that much of the resulting benefit goes unmeasured’.

The existence of this quality-improvement problem, broadly interpreted by Aghion and Howitt (1998) to encompass both actual quality improvements and new goods, implies that an accelerating rate of technological change may be largely expressed in (mounting) quality improvements and the (faster) introduction of new products rather than in a rising rate of per-capita output growth. This failure of (conventional) measures of output to capture technological change could be the reason why the national income and product statistics do not show rising growth rates despite persistent size increases (see also Dinopoulos and Thompson, 1999: 166).

In its core, our analysis consists of a simple test to see if this suggested explanation can indeed account for non-accelerating growth and the absence of scale effects in empirical studies so far. For this purpose, we build on literature that has examined the role of measurement problems in a similar growth puzzle, namely the post-1973 ‘productivity slowdown’ (e.g. Baily and Gordon, 1988; Diewert and Fox, 1999; Sichel, 1997). This literature finds that the existence of an upward bias in measured prices, in particular the consumer price index (CPI), leads to undermeasurement of the level and growth of real output, which may account for the marked slow-down of productivity growth after 1973. More precisely, the idea is that mismeasurement, and especially the part due to the quality-improvement problem, is particularly severe in sectors that have gradually become more important, and that the growing share of unmeasurable sectors in the total economy could explain the slowdown—or, alternatively, non-accelerating growth.



In the next section, we discuss some problems with quantifying growth and the measurement of prices, specifically the quality-improvement problem, in more detail. We build on analyses of the productivity slowdown to establish how such measurement difficulties can provide a solution to the scale effects problem.

### **3. Bias in the CPI and Increasing Undermeasurement of Growth**

Because ‘much of the growth of productivity and output in the long-run is the result of product innovations that generate new and improved goods whose contribution to output are only partially measured’ (Aghion and Howitt, 1998: 439), one should exercise some caution in drawing definite conclusions concerning the acceleration of output growth using ‘plain’ data on GDP or a related statistic. Unfortunately, prices, e.g. the CPI, are also among the most difficult variables to quantify. There has been—and still is—a lot of concern about the accuracy with which prices are measured. In the U.S., this concern has led to the establishment of the Boskin Commission (the Advisory Commission to study the Consumer Price Index, chaired by Michael Boskin) several years ago. Its main task was to investigate the accuracy with which the CPI measures the real change in the cost of living. The Commission has found an upward bias: the CPI overstates the change in the actual cost of living by about 1.1% per year (Boskin et al., 1996).

Because prices are so important, the observed inaccuracy of the CPI in quantifying the actual cost of living has far-reaching consequences. It affects, for instance, the conduct of monetary policy, but of most interest in the context of this paper, is that it has an impact on determining the rate at which an economy grows (Eldridge, 1999; Landefeld and Grimm, 2000). A rough estimate is that the upward bias in the CPI, through mismeasurement of consumption (making up about two-thirds of GDP), causes real GDP growth to be understated by a half percentage point per year (Boskin and Jorgenson, 1997: 92). Likewise, the 1.1% upward bias in the CPI corresponds with a 1.1% downward bias in real wage growth.

The Boskin Commission discerns several sources of bias in the CPI: substitution bias, outlet substitution bias, quality change bias, new product bias and formula bias. Of these, the bias resulting from quality change in existing products and the introduction of

new products has received most attention. Because the quality-improvement problem is centred on quality change and new products as well, these types of bias are of most interest to us; the other biases are not directly related to R&D and technological progress, i.e. to endogenous growth theory.

Quality change and new product bias, or better, the quality-improvement problem, can be addressed, at least partly, by using hedonic pricing methods. These methods construct the price of a certain product from the price of a reference product, to which the prices paid for different quality characteristics are added or subtracted (cf. Triplett, 1987). The Boskin Commission relies on other studies applying these techniques for its assessment of the bias resulting from quality change and new products for several subcomponents of the CPI (Boskin et al., 1996). Summing the weighted contributions of the bias in these subcomponents, it estimates that the size of this type of bias in the CPI equals 0.60%, which is some 55 percent of the total bias in the CPI (1.1%). The quality-improvement problem is a serious problem then: it translates into the understatement of real GDP growth by some 0.25-0.3% per year (55 percent of 0.5%).

The findings of the Boskin Commission bear great relevance to the proper calculation of growth rates. Obviously, the positive trend in real output is understated due to the quality-improvement problem and the corresponding upward bias in measured prices. Yet, this downward bias in real GDP growth does not necessarily entail that growth actually does accelerate and exhibits a scale effect. Baily and Gordon (1988) suggest two ways in which the upward bias in price indices may result in rising undermeasurement of the level and growth of real GDP: (1) the upward bias increases over time, and (2) the share in total production of products and sectors plagued by understating measurement errors grows.

The possibility of an increasing upward bias has been explored by the Boskin Commission itself. For several of the 27 subcomponents of the CPI, the Boskin Commission has assessed the quality change and new product bias over different time intervals, and Table 1 summarizes the resulting estimates.

[Insert Table 1 about here]

The table shows that the bias in those components weighing heavily in the CPI, e.g. Apparel and upkeep, New vehicles, and Motor fuel, has risen (in bold). This would suggest that the overall bias in the CPI has increased over time. And in fact, the Commission concludes that in 1980 quality change and new product bias equalled 0.488% per year, while it equalled 0.613% in 1996 (Boskin et al., 1996). Work by Costa (2001) and Hamilton (2001), however, contradicts this. Using Engel curves, Costa (2001) estimates that in the 1960s the CPI bias was 0.4%, after which it rose to 2.7% between 1972 and 1982, only to decline again to 0.6% in the period 1982-1994. Hamilton (2001) applies a similar technique and finds that it equalled roughly 2.5% per year between 1974 and 1981, but slightly under one percent per year from 1981 to 1991. By and large, it seems unlikely that since the 1950s total bias in measured prices has increased steadily.

Regarding the possibility of a fundamental relation between mismeasurement and sectoral composition, empirical evidence is non-existing to our knowledge. It is however plausible that the share of sectors subjected to undermeasurement has increased over time. Numerous scholars have examined whether the combined effect of a changing sectoral structure and upward bias in prices can account for the so-called productivity slowdown (e.g. Baily and Gordon, 1988; Diewert and Fox, 1999; Sichel, 1997). The idea is that services make up an ever-larger share of the total economy, and that in this sector output is most difficult to measure and overall understated. That is to say, the sectors in which the problems of quantification are particularly severe have gradually become more important.<sup>2</sup>

Griliches (1994: 10) goes one step further and divides sectors in reasonably measurable (e.g. wheat production) and the rest (e.g. lawyer services; unmeasurable). He subsequently observes that, 'in the early post-World War II period, the situation was not all that bad: about half of the overall economy was 'measurable' in this sense. By 1990, however, the fraction of the economy for which the productivity numbers are half reasonable had fallen to below one-third'. If unmeasurable effectively involves the

---

<sup>2</sup> This holds most for the quality-improvement problem / quality change and new product bias; the other sources of bias in the CPI are much less concentrated in certain sectors of the economy.

understatement of actual production in these sectors, as Griliches assumes, undermeasurement of both the level and rate of growth of real output and income has accumulated, at least between the early post-World War II years and 1990.

In the next section we use corrected data for the unmeasurable sectors to see whether the quality-improvement problem, or in fact mismeasurement of prices in general, in combination with structural change, can account for non-accelerating growth.

#### **4. Testing for accelerating growth with corrected data**

The combined effect of an economy increasingly producing goods in unmeasurable sectors and an upward bias in prices implies a rising undermeasurement of real GDP and real GDP growth. It is thus legitimate to ask whether measurement difficulties and a changing sectoral structure together hold the key to solving the scale effects problem. To investigate this matter we first correct the data on output and growth for undermeasurement concentrated in certain sectors. We then repeat some of the tests applied in previous empirical work on growth scale effects—and that of Jones (1995a) in particular—using the corrected data.

##### **4.1 Data**

Because, first, an observed acceleration in the rate of increase of per-capita GDP might just as well be due to the acceleration of growth of resources used as productive inputs, for example labour, and, second, the relation between the stock of knowledge and per-capita GDP growth is not one-to-one, evidence of growth scale effects should best not be sought in GDP growth trends but rather in the trend of labour productivity (per hour) or even better TFP.

Therefore we use data on i) gross value added at factor cost, ii) number of employees and iii) hours worked in a sample of ten sectors and nine countries. Data are from van Ark (1996) and are available electronically at [http://www.ggdc.net/databases/10\\_sector95.htm](http://www.ggdc.net/databases/10_sector95.htm). Our data span the period 1950-1997.

The database allows us to calculate productivity (per hour worked or per worker) for each of the sectors, and to trace its development over 48 years.

The sectors are Agriculture, Mining, Manufacturing, Public Utilities, Construction, Wholesale and Retail Trade, Transport and Communication, Finance, Insurance and Real Estate, Community, Social and Personal Services, and Government Services. The countries are Denmark, France, West Germany, Italy, Japan, Spain, Sweden, the United Kingdom, and the United States.

Following Griliches (1994: 11) we discern five measurable sectors: Agriculture (Ag), Mining (Mi), Manufacturing (Ma), Public Utilities (PU), and Transport and Communication (TaC) (with the note that the latter two categories are combined in Griliches 1994 and referred to as transportation and utilities). Similarly, five sectors are considered unmeasurable: Construction, Wholesale and Retail Trade, Finance, Insurance and Real Estate, Community, Social and Personal Services, and Government Services (with the note that in Griliches 1994 Wholesale and Retail Trade, represent two sectors, referred to as wholesale trade and retail trade respectively).

## 4.2 Analysis and Results

Following Baily and Gordon (1988), we look for growth scale effects in measurable sectors (Ag, Mi, Ma, PU and TaC) in the United States. As in Jones (1995a), we test for a time trend and apply an augmented Dickey-Fuller (ADF) test. This test allows us to determine whether a unit root exists in a time series model (Dickey and Fuller, 1979; Greene, 2003). Table 2 presents the results.

[Insert Table 2 about here]

Both tests provide strong evidence that over the period 1950-1997 growth of real GDP per hour worked has not increased. Generally, the time trend is negative rather than positive, though only for Public Utilities we obtain a statistically significant relationship. We can robustly reject the null hypothesis of a unit root, implying there is non-stationarity in growth rates in all measurable sectors.

Still, undermeasurement of growth in certain sectors may be causing the absence of a scale effect in aggregate-level growth. We repeat our tests at the aggregate level using corrected data for the unmeasurable sectors. The idea is to correct for possible downward biases in the trend of real value added per hour in the unmeasurable sectors by simply adding a chosen number of percentage points to observed growth in these sectors. Aggregate-level growth is then straightforwardly calculated by summing the (corrected) sectoral growth rates, weighted by each sector's share in the total economy. This correction will manifest itself in aggregate-level growth in two ways. Firstly, and most logically, it will result in a higher *level* of growth. Secondly, as we discussed in the previous section in some detail, it will change the *pattern* of growth: through the changing sectoral structure of the economy, correcting for undermeasurement in the unmeasurable sectors manifests itself in the development over time of aggregate-level productivity growth. If the combined effect of structural change and measurement errors is large enough, this will result in a pattern consistent with the scale effects prediction of endogenous growth models.

The complete method for the construction of aggregate-level growth rates corrected for measurement problems and sectorally concentrated undermeasurement is as follows:

- first, the percentage increase of labour productivity per hour for a sector over time  $t$  and time  $t + 1$  is calculated (log-linear growth rates);
- next, this sectoral growth rate, with or without a correction for undermeasurement applied, is multiplied by the sector's average share in total hours over time  $t$  and time  $t + 1$ . We obtain a weighted sectoral rate; and
- finally, the weighted sectoral growth rates—remember that there are ten sectors, half of which are measurable while the other half is unmeasurable—are summed to obtain the (corrected) aggregate-level rate at which aggregate productivity grows.

Because we simply average shares over time  $t$  and time  $t + 1$ , the calculated growth rates of output per hour in the U.S. with no correction (i.e. 0%) do not equal the originally observed aggregate-level rates. The original rate of growth of value added per hour and the growth rate of value added per hour based on a zero correction, are nevertheless strongly correlated ( $\rho = 0.943$ ). And although the use of average shares invokes an

upward bias in (uncorrected) aggregate-level growth, this bias is not statistically significant at usual levels.

Figure 2 depicts growth of labour productivity per hour constructed using six different corrections, 0, 1, 5, 10, 15 and 20% respectively, for the unmeasurable sectors. As noted above, adding (a) percentage point(s) to correct for a possible downward bias in measured growth in the unmeasurable sectors leads to larger aggregate-level growth rates. Moreover, though again as expected, the development of ‘observed’ productivity growth changes with the size of the correction.

[Insert Figure 2 about here]

Most importantly, Figure 2 presents some additional insights into whether the upward bias in prices and the associated undermeasurement of the upward trend in real output can explain the absence of growth scale effects. It appears they do: with high corrections applied, i.e. adding more than 10% growth of labour productivity per hour in the unmeasurable sectors, the graph suggests that the aggregate-level growth rate could very well increase over time.

Further evidence comes from the formal analysis. In Table 3 we show that the existence of a unit root in the growth rate of value added per hour in the United States is strongly rejected, regardless of the correction applied. The time-trend test discloses that with the application of a 10% correction, this rate does not yet exhibit a scale effect. Only when we raise measured growth in the unmeasurable sectors with more than 10% does the aggregate-level rate exhibit a statistically significant positive time trend, that is to say, does it accelerate over time. Interestingly, correcting for undermeasurement has different effects in the time trend test than in the ADF test. With increasing corrections, the time trend shows the expected gradual movement from a negative to a positive time trend, whereas the hypothesis of a unit root is more strongly rejected up till a 10% correction. Only when larger corrections (15%, 20%) are applied does the likelihood of non-stationary growth rates increase.

[Insert Table 3 about here]

This 10% correction does not compare favourably to the Boskin Commission's estimate of 1.1% upward bias in the CPI nor even to some of the highest estimates of 2.5-2.7% bias (Costa, 2001; Hamilton, 2001). Therefore, we conclude that measurement problems such as the quality-improvement problem or other sources of bias in measured prices do not solve the scale effects problem: there is no acceleration of growth in the productivity statistics.<sup>3</sup>

#### 4.3 Robustness

We test the robustness of this finding across different dimensions. First, we include eight additional countries in our analysis. The countries we add are Denmark, France, West Germany, Italy, Japan, Spain, Sweden, and the United Kingdom. Combined, these countries account for the majority of worldwide R&D efforts. This means that we can account for the problem that we may not find country differences in the rate at which output per capita increases because knowledge rapidly diffuses across countries. As Jones (1995a: 519) puts it, 'to the extent that technology diffuses quickly across international boundaries, testing the R&D equation country-by-country may produce misleading results. Perhaps the correct unit of analysis is the entire OECD or even the world instead of an individual country'.

We apply the main tests from the previous section to a hypothetical world aggregate, constructed by weighting the contribution of the nine countries according to their share in total employment (averaged over time  $t$  and time  $t + 1$ ). Within this aggregate, we again correct the unmeasurable sectors for downward bias due to mismeasurement of prices. Because we use data on output per worker, we adjust the

---

<sup>3</sup> Obviously, the correction applied does not have an exact counterpart in CPI bias. They are in the 'same league', however, and we can meaningfully compare them. In addition, upward bias in the CPI bias does have a counterpart in the understatement of real wage growth; and, in the long run, productivity growth is not just the main but also the only source of real wage growth.



sectoral weighting procedure correspondingly. The ten sectors contribute to the aggregate according to their share in total employment (rather than hours worked), where shares are simply averaged over time  $t$  and time  $t + 1$ . Hence, again the calculated growth rates of output per worker with no correction for undermeasurement applied do not exactly equal the original aggregate-level rates.

The results of the analysis at the world level closely resemble those found for the United States (see Table 4).

[Insert Table 4 about here]

Undermeasurement does not account for the absence of growth scale effects. The ADF test robustly rejects the hypothesis that growth of output per worker has a unit root. In addition, the correction required for the growth rate to exhibit a significant, positive time trend, in this case almost 20%, does not compare favourably to actually observed (quality change and new product) bias in the CPI.<sup>4</sup>

Four other aspects of the analysis presented in the previous subsection that deserve further scrutiny are the possible role played by the 1973 slowdown, the robustness of the results to different classification of sectors as measurable and unmeasurable, and the sensitivity of the findings to different measures of scale and to different weighting procedures. We find that none of these affects the conclusion that measurement difficulties cannot account for the absence of an accelerating trend in growth (results not shown, but available upon request).

First, testing for growth scale effects in the pre- and post-1973 periods separately has a negligible impact on the results. The evidence still indicates that growth has a positive time trend only at implausibly high levels of undermeasurement, and the hypothesis of a unit root is strongly rejected at all levels of correction. Second, because all the measurable sectors have decreasing hours shares and all the unmeasurable sectors

---

<sup>4</sup> This extended dataset further enables us to perform cross-sectional tests along the lines of Backus et al. (1992). We find that growth of output per worker does not increase with sector size. This result is robust to the inclusion of different sets of control variables.

have increasing hours shares, the results are fully robust to different measurable-unmeasurable classifications. Third, checking for a trend in models with employment, total hours and total value added respectively (all taken from van Ark 1996) as independent size variables, confirms that growth does not increase with scale. Finally, the possibility that the use of an employment-based weighting procedure rather than an output-based one drives our conclusion is also dismissed easily: as a share in total value added, the unmeasurable sectors have increased from about 58% of the economy in 1950 to about 62% of the economy in 1997. Meanwhile, the hours share of these sectors has increased steadily from roughly 54% in 1950 to 75% in 1997. Therefore, if anything, undermeasurement due to measuring problems is more likely to solve the scale effects problem when our weighting procedure is used than when an output-based one is used.<sup>5</sup>

---

<sup>5</sup> Note that summing weighted sectoral growth rates to obtain a (corrected) growth rate at the aggregate level is not inappropriate. Generally, national accounts are first constructed at the aggregate level and only in second instance are they disaggregated and is output or value added assigned to different sectors (cf. Baily and Gordon, 1988; Sichel, 1997).

Since the aggregate is fixed in this respect, the idea that real output and growth thereof is understated in some sectors has the implication that they are overstated in other sectors. It should be clear that our method of correcting growth of labour productivity per hour for undermeasurement does not assume such ‘reshuffling’ of output and growth (Baily and Gordon, 1988: 349). Simply adding a chosen level of correction to real growth in the unmeasurable sectors is the same as assuming that real GDP and increases of real GDP have been understated. That is, our interpretation of the quality-improvement and related measuring problems has been that it leads to unmeasured ‘benefits’ (Aghion and Howitt, 1998: 438), and understatement of output per capita and the rate at which it increases. Strict adherence to the idea of a fixed GDP, on the other hand, precludes the possibility that problems of measurement in combination with sectoral change and unmeasurable sectors have any bearing on the issue whether or not growth exhibits a scale effect and accelerates.

Overall, the robustness tests strengthen our conclusion that it is implausible that an upward bias in measured prices (in particular the CPI) accounts for the absence of growth scale effects.

## **5. Concluding Remarks**

Early models of endogenous growth consistently predict the existence of growth scale effects: larger economies grow faster than smaller economies do, which amounts to an accelerating trend in output. The empirical literature on growth scale effects has found no support for this theoretical prediction. These empirical tests, however, have not taken into account that mismeasurement of the growth variables may be a cause for the impossibility to empirically establish accelerating growth. This is surprising because Aghion and Howitt (1998) already suggested that especially the quality-improvement problem might drive the results obtained in empirical work. Price indices such as the CPI are biased upward because they do not accurately measure the benefits of new and improved products. This, in turn, leads to undermeasurement of real output and productivity growth.

The relevance of this measurement problem for explaining the growth-scale effect has been neglected and its impact is not yet systematically explored. Inspired by parallel literature relating mismeasurement of prices to the post-1973 productivity slowdown, we have tested in this paper if measurement problems (such as the quality-improvement problem), in combination with a changing sectoral structure, ‘solve’ the scale effects problem and lead to the establishment of accelerating growth patterns. Our analysis convincingly shows that they do not. Firstly, we have dismissed the possibility that the upward bias in prices has worsened. Secondly, using corrected data on real GDP per hour worked we find that growth only accelerates at implausibly high levels of bias in measured prices. Upper bounds on estimates for total CPI bias are in the range of 2.5-2.7%, whereas the pace at which productivity increases only exhibits a scale effect if this bias is 10% or more.

Because of this measurement-proof empirical refutation of the growth-scale effect, the most logical direction for future work aimed at getting a better grasp of this

problem lays in the domain of improved theory. There are two ways out of this. The first simply entails the development of models of endogenous growth without scale effects but, so far, this has been rather unsuccessful. Models of growth without scale effects only yield endogenous long-run growth under very strict linearity assumptions. Hence, the second solution, which has been suggested by Jones (2003). According to him, an important challenge for endogenous growth theory is to provide ‘an intuitive and compelling justification’ for its crucial, knife-edge assumptions (Jones, 2003: 499). A particularly interesting attempt at providing such justification subsequently is by Peretto and Smulders (2002). They use the concept of technological distance—much used in empirical studies on R&D and knowledge spillovers—as the microfoundation for endogenous growth without scale effects. Although, ultimately, Peretto and Smulders’ (2002) results appear driven by the unrealistic assumption that firms randomly choose their R&D activities, their work deserves follow up.

At the same time, we should not dismiss an empirical explanation altogether. Nordhaus’ (1997) work on increases in lighting efficiency from the early days of Peking man to present day supports the accelerating growth prediction of endogenous growth models. In particular, he finds that the rate of improvement in lighting efficiency has increased over time. This triggers a more fundamental question regarding the nature of growth. Does the conceptualisation of growth in models of endogenous growth actually fit standard measures of growth based on aggregates from the System of National Accounts? What might be needed is a whole new approach to measuring growth.

The fact that both empirical studies (including ours) and revised theoretical models of growth without scale effects have not fully solved the scale effects problem suggests that the problem might be more fundamental than realised. As such, the scale effects problem deserves further exploration.

## References

- Aghion, P., Howitt, P., 1992. A model of growth through creative destruction. *Econometrica*, 60(2), 323-351.

- Aghion, P., Howitt, P., 1998. *Endogenous growth theory*. Cambridge MA: The MIT Press.
- Backus, D.K., Kehoe, P.J., Kehoe, T.J., 1992. In Search of Scale Effects in Trade and Growth. *J. of Econ. Theory*. 58(2), 377-409.
- Baily, M.N., Gordon, R.J., 1988. The Productivity Slowdown, Measurement Issues, and the Explosion of Computer Power. *Brookings Pap. on Econ. Activity*. 1988(2), 347-420.
- Boskin, M.J., Dulberger, E.R., Gordon, R.J., Griliches, Z., Jorgenson, D.W., 1996. Toward A More Accurate Measure Of The Cost Of Living. The final report from the 'Boskin Commission' to the Senate Finance Committee. <http://www.socialsecurity.gov/history/reports/boskinrpt.html>. Accessed 25 March 2009.
- Boskin, M.J., Jorgenson, D.W., 1997. Implications of Overstating Inflation for Indexing Government Programs and Understanding Economic Progress. *Am. Econ. Rev. Pap. and Proc.* 87(2), 89-93.
- Costa, D.L., 2001. Estimating Real Income in the United States from 1888 to 1994: Correcting CPI Bias Using Engel Curves. *J. Political Econ.* 109(6), 1288-1310.
- Dickey, DA, Fuller, W.A., 1979. Distribution of the Estimators for Autoregressive Time Series with a Unit Root. *J. of the Am. Stat. Assoc.* 74(June), 427-431.
- Diewert, E.W., Fox, K.J., 1999. Can measurement error explain the productivity paradox? *Can. J. of Econ.* 32(2), 251-280.
- Dinopoulos, E., Thompson, P., 1999. Scale effects in Schumpeterian models of economic growth. *J. of Evol. Econ.* 9(2), 157-185.
- Eldridge, L.P., 1999. How price indexes affect BLS productivity measures. *Mon. Labor Rev.* February, 35-46.
- Greene, W.H., 2003. *Econometric Analysis*. Fifth Edition. New Jersey: Prentice Hall.
- Griliches, Z., 1994. Productivity, R&D, and the Data Constraint. *Am. Econ. Rev.* 84(1), 1-23.
- Hamilton, B.W., 2001. Using Engel's Law to Estimate CPI bias. *Am. Econ. Rev.* 91(3), 619-630.

- Jones, C.I., 1995a. Time series tests of endogenous growth models. *Q. J. of Econ.* 110(May), 495-525.
- Jones, C.I., 1995b. R&D-Based Models of Economic Growth. *J. Political Econ.* 103(4), 759-784.
- Jones, C.I., 2003. Population and Ideas: A Theory of Endogenous Growth, in: P. Aghion, R. Frydman, J. Stiglitz, M. Woodford (Eds.), *Knowledge, Information, and Expectations in modern Macroeconomics: In Honor of Edmund S. Phelps*. Princeton / Oxford: Princeton University Press, pp. 498-521.
- Landefeld, J.S., Grimm, B.T., 2000. A note on the impact of hedonics and computers on real GDP. *Surv. of Curr. Bus.* 80(12), 17-22.
- Li, C.-W., 2000. Endogenous vs. semi-endogenous growth in a two-R&D-sector model. *Econ. J.* 110(March), C109-122.
- Li, C.-W., 2002. Growth and scale effects: the role of knowledge spillovers. *Econ. Lett.* 74(2), 177-185.
- Lucas, R., 1988. On the mechanics of economic development. *J. Monetary Econ.* 22(1), 3-42.
- Nordhaus, W.D., 1997. Do Real Output and Real Wage Measures Capture Reality? The History of Lighting Suggests Not, in: T.F. Bresnahan, R.J. Gordon (Eds.), *The Economics of New Goods*. Chicago: The University of Chicago Press, pp. 27-70.
- Peretto P.F., Smulders, S., 2002. Technological distance growth and scale effects. *Econ. J.* 112(July), 603-624.
- Petty, W., 1682. *Essays on Political Arithmetic*. <http://www.gutenberg.org/etext/5619>. Accessed 25 March 2009.
- Romer, P.M., 1986. Increasing returns and long-run growth. *J. Political Econ.* 94(5), 1002-1037.
- Romer, P.M. (1990). Endogenous Technological Change. *J. Political Econ.* 98(5), S71-102.
- Romer, P.M., 1994. The Origins of Endogenous Growth. *J. of Econ. Perspect.* 8(1), 3-22.
- Segerstrom, P., 1998. Endogenous Growth Without Scale Effects. *Am. Econ. Rev.* 88(5), 1290-1310.

- Sichel, D.E., 1997. The productivity slowdown: Is a growing unmeasurable sector the culprit? *Rev. of Econ. and Stat.* 79(3), 367-370.
- Smulders, S., van de Klundert, T., 1995. Imperfect competition, concentration and growth with firm-specific R&D. *Eur. Econ. Rev.* 39(1), 139-160.
- Triplett, J., 1987. hedonic functions and hedonic indexes, in: J. Eatwell, M. Milgate, P. Newman (Eds.), *The New Palgrave: A Dictionary of Economics*. Vol. 2. London: MacMillan, pp. 630-634.
- van Ark, B., 1996. Sectoral growth accounting and structural change in post-war Europe, in: B. van Ark, N. Crafts (Eds.), *Quantitative aspects of post-war European economic growth*. Cambridge: Cambridge University Press, pp. 84-164.
- Young, A., 1998. Growth without Scale Effects. *Journal of Political Economy*, 106(1), 41-63.

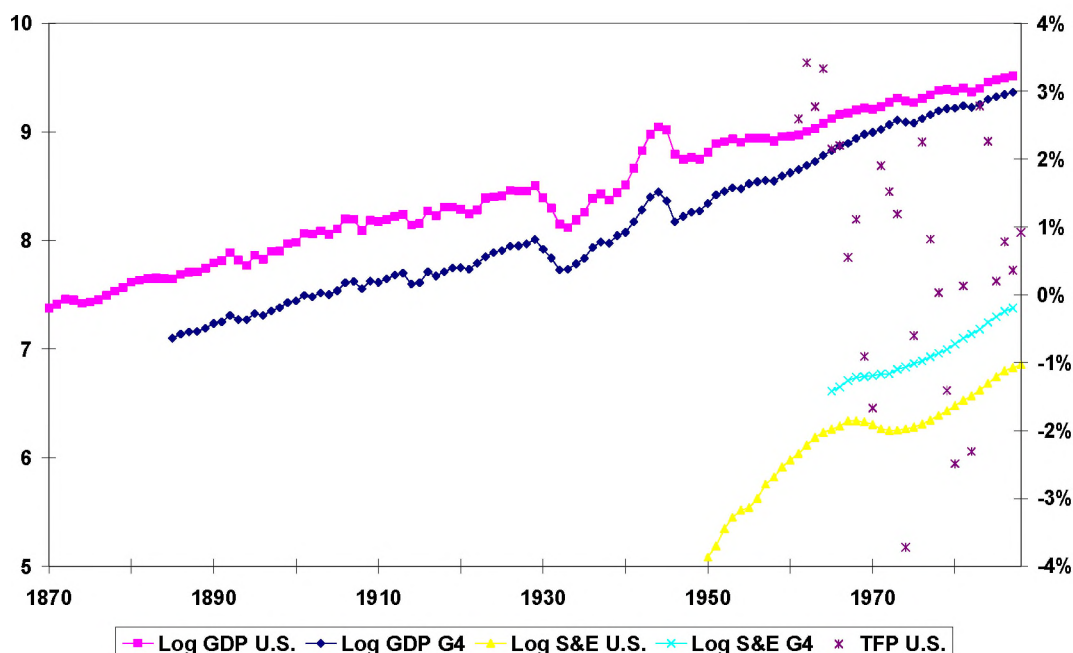


Figure 1, Natural logarithm of GDP per capita 1870-1987 (GDP), and scientists and engineers engaged in R&D 1950-1988 (S&E), for the United States (U.S.) and G4 (France, Germany, Japan and the U.S.) (left axis). Total factor productivity growth for the U.S. 1961-1988 (TFP U.S.) (right axis). Source: Jones (1995a), and own calculations.

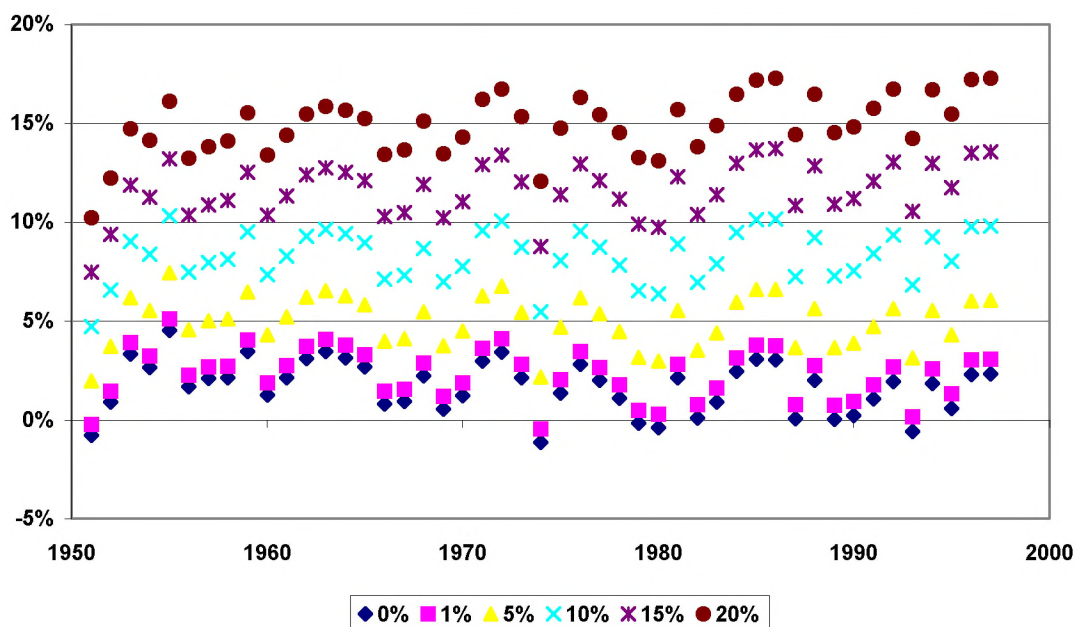


Figure 2, Growth rate of value added per hour in the U.S., 1951-1997, with different corrections for the unmeasurable sectors applied. Source: van Ark (1996), and own calculations.



Table 1: Quality change and new product bias for some CPI (sub)components.

CPI (sub)component	Estimated annual bias	Time interval	Relative importance (%)
Appliances, including electronic (Housing)	<b>3.6</b> <b>5.6</b>	<b>1973-1994</b> <b>1994-1996</b>	.806
(Apparel and upkeep)	<b>-0.95</b> <b>1.00</b>	<b>1965-1985</b> <b>1985-1996</b>	5.516
New vehicles (Transportation)	<b>0.00</b> <b>0.59</b>	<b>1970-1983</b> <b>1983-1996</b>	5.027
Used cars (Transportation)	2.44 1.59	1967-1987 1987-1996	1.342
Motor fuel (Transportation)	<b>0.0</b> <b>0.25</b>	<b>1974-1984</b> <b>1984-1996</b>	2.908
Public transportation (Transportation)	<b>2.66</b> <b>4.60</b> 0.00	<b>1972-1977</b> <b>1977-1982</b> 1982-1996	1.523
Prescription drugs (Medical care)	3.00 2.00	1970-1995 1995-1996	.891

Remarks: Major components in parenthesis (e.g. Housing). Source: Boskin et al. (1996).

Table 2: Time-trend and ADF tests of productivity growth rates for the measurable sectors.

Sector	Time trend	ADF test
Agriculture (Ag)	-2.57 (7.83)	-1.21 (.15)***
Mining (Mi)	-5.95 (7.55)	-.56 (.13)***
Manufacturing (Ma)	1.52 (2.38)	-.79 (.15)***
Public Utilities (PU)	-10.4 (4.45)**	-.65 (.14)***
Transport and communication (TaC)	-3.43 (2.91)	-1.32 (.19)***

Remarks: The time trend test is performed by estimating the equation  $g_t = \beta_0 + \beta_1 \text{year}/10,000 + \varepsilon_t$ . The table reports estimates for  $\beta_1$ . For the ADF test the equation  $\Delta g_t = \alpha * g_{t-1} + \beta + \gamma(LL) * \Delta g_{t-1} + \varepsilon_t$  is estimated, where  $\gamma(LL)$  is the 'lag length' chosen using the Schwartz information criterion (cf. Jones 1995a). The table reports values for  $\alpha$ . The null hypothesis of  $\alpha = 0$  (there is a unit root in growth rates) is evaluated against the alternative,  $\alpha < 0$  (there is no unit root in growth rates). (\*), (\*\*), and (\*\*\*) denotes significance at the 10%, 5% and 1% level respectively. Standard errors in parentheses. Critical t-values for the ADF-test are: -2.60 for 10%, -2.93 for 5% and -3.58 for 1%.

Table 3: Time-trend and ADF tests for different percentage points correction.

Correction	Time trend	ADF test
0%	-2.68* (1.37)	-.76 (.14)***
1%	-2.26 (1.37)	-.78 (.14)***
5%	-.61 (1.38)	-.83 (.14)***
10%	1.46 (1.39)	-.84 (.14)***
15%	3.52** (1.41)	-.78 (.13)***
20%	5.59*** (1.42)	-.68 (.13)***
See notes to Table 2.		

Table 4: Time-trend and ADF tests for corrected world-level growth rates.

Correction	Time trend	ADF test
0%	-9.86*** (1.70)	-.44 (.14)**
1%	-9.17*** (1.70)	-.46 (.15)**
5%	-6.44*** (1.70)	-.60 (.16)***
10%	-3.01* (1.70)	-.78 (.18)***
15%	.41 (1.69)	-.86 (.18)***
20%	3.83** (1.69)	-.76 (.18)***

See notes to Table 2. Critical t-values for the ADF-test are: -2.62 for 10%, -2.95 for 5% and -3.65 for 1%.